DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health

Division of Research Resources
Biomedical Research Technology Program
Annual Progress Report
PART I, TITLE PAGE

1. PHS AWARD NUMBER: 1 U 4 1 R R 0 1 6 8 5 - 0 2

2. TITLE OF AWARD BIONET, National Computer Resource for Molecular Biology

3. NAME OF RECIPIENT INSTITUTION: IntelliGenetics, a subsidiary of IntelliCorp

4. HEALTH PROFESSIONAL SCHOOL (If applicable): N/A

5. REPORTING PERIOD:

5a. FROM (Month, Day, Year): 0 3 - 0 1 - 8 5

5b. TO (Month, Day, Year): 0 2 - 2 8 - 8 6

6. PRINCIPAL INVESTIGATOR:

6a. NAME: Ralph E. Kromer

6b. TITLE: President, IntelliCorp

6c. SIGNATURE: Ralph E Kromer

7. DATE SIGNED (Month, Day, Year):

8. TELEPHONE (Include Area Code): 4 1 5 - 9 6 5 - 5 7 5 8

# Part II. Description of Research Progress

This section of the Report provides statistical information on use of the BIONET$^{tm}$ Resource. The period covered is 12/84 - 11/85. Its individual sections have been prepared under guidelines discussed with BRTP staff. In the one year since BIONET formally began accepting and approving applications for access, the BIONET community has expanded to 533 Principal Investigators (PI's), representing about 1800 total investigators, whose classification is as summarized in Table II-1. As shown in the Table, of the 537 accepted Class I PI's, 27 chose to drop their membership when the subscription fee for telecommunication access was announced. To review, Class I PI's represent the service component of the Resource, Class II the collaborative component, and Class III is reserved for those persons who are responsible for local computing facilities, and who have agreed to act to support the local community of scientists accessing BIONET.

**Table II-1:** BIONET User Community

| | |
|---|---|
| Class I PI's | 537 |
| Declined | (27) |
| Total Class I | 510 |
| Class II PI's | 12 |
| Class III PI's | 3 |
| Nat. Advis. Comm. | 8 |
| Net Total | 533 |

# II.A. Scientific Subprojects

### II.A.1. Collaborative Research and Service

We report Collaborative Research and Service first, because it represents the bulk of the Resource's activities in this grant year. We have not reported the research abstracts in the interests of saving space. These abstracts are maintained on our PC database of BIONET users, and are available to any interested party.

We report the "Usage Factor" as both central processor units (cpu time), in minutes, and connect time, in hours, for each PI. These values are the sums of all usage by the PI and his or her group members ("Sub-I's"). We report data only on those PI groups that have accessed the Resource. Of the 522 Class I and II PI's, 361, representing about 1010 individual investigators, have logged on to the Resource. The data for Sub-I's is maintained on our PC database and is available to any interested party.

We do not report Resource staff hours nor BRTP funds allocated for individual PI's because it is impossible to allocate these rationally to such a large user community. Summary information for all PI's is given in Part II, Section C., the Resource Summary Table.

### II.A.2. Core Research and Development

We report summary information for our two current Core Research projects. The BRTP funds allocated are calculated on the basis of the sum of individual salaries for the time spent, plus a percentage of the total Resource budget devoted to providing supporting services.

The Resource's primary commodity is cpu cycles made available to all investigators, including BIONET staff. Thus, the percentage used in the above calculation is the cpu time used by the indicated investigators compared to the total cpu time used on the system. The calculation for the actual cost is this percentage times the basis of costs of providing this commodity to local staff; i.e., the budget categories of Supplies, Computer Expenses, Software Licenses and Services and Documentation. This basis is, using our anticipate expenditures in these categories through February 28, 1986, $265,500.

# DRR SCIENTIFIC SUBPROJECT FORM

**PART II, SECTION A** — **INSTITUTION:** IntelliGenetics

**AWARD NUMBER** 1 U 4 1 R R 0 1 6 8 5 — 0 2 — **REPORT PERIOD:** March 1, 1985 **to** February 28, 1986

**Fill out a separate Subproject Form for Core, Collaborative or Training. Check one of the following:**

[XXXX] **CORE RESEARCH & DEVELOPMENT** [ ] **COLLABORATIVE RESEARCH & SERVICE** [ ] **TRAINING**

| **Descriptive Title (80 characters)** (1)<br>**Abstract** | (2) Science Code Axis I | (2) Science Code Axis II | (3) a. Investigator(s) Name (Last Name, First Name & Middle Initial)<br>b. Department<br>c. Non-Host Inst. | (4) USAGE FACTOR: Resource Technology a/ | (4) USAGE FACTOR: CPU MIN USED | (4) USAGE FACTOR: Resource Staff Hours | (5) BRTP Fund Allocat |
|---|---|---|---|---|---|---|---|
| BIONET<br>Satellite Program<br><br>Distribution of the computational and communication facilities of BIONET to other, geographically remote sites (Satellite Resources); linking these Satellites for rapid exchange of electronic mail and bulletins, data files and programs | 9 | 40, 42 | a. Smith, Dennis H.<br>Roode, R. David<br><br>b. IntelliGenetics | DEC 2060<br>" | 22<br>284 | 94<br>148 | $ 11,3 |
| Sequential Search Machine<br><br>Exploring the new technologies of hardware, sequential search machines for applications to biological sequence retrieval and intercomparison | 9 | 42, 68<br>70 | a. Smith, Dennis H.<br>Brutlag, D.<br><br>b. IntelliGenetics | DEC 2060<br>" | 45<br>276 | 188<br>40 | $ 16,20 |
| **CUMULATIVE TOTALS:** **No. subprojects** | 2 | | | | 627 | 470 | $ 27,54 |

a/ Identify Resource Technologies Used. b/ Give Hours Resource Technologies Used. See Instructions, page ii.

### II.A.3. Training

We report summary information for our Training program. The sites at which BIONET provided some level of training are named here, and are discussed in more detail in Part III, the Narrative Description. The BRTP funds allocated include the salaries of the personnel involved, the estimated costs of the training program itself for this year ($5,500), and a percentage allocation of cost for cpu use, as discussed in the previous section.

# DRR SCIENTIFIC SUBPROJECT FORM

PART II, SECTION A

INSTITUTION: Intelligenetics

AWARD NUMBER: 1 U 4 1 R R 0 1 6 8 5 — 0 2

REPORT PERIOD: March 1, 1985 to February 28, 1985

Fill out a separate Subproject Form for Core, Collaborative or Training. Check one of the following:

☐ CORE RESEARCH & DEVELOPMENT ☐ COLLABORATIVE RESEARCH & SERVICE ☒ XXXX TRAINING

| Descriptive Title (80 characters) (1)<br>Abstract | (2) Science Code Axis I | (2) Science Code Axis II | (3) a. Investigator(s) Name (Last Name, First Name & Middle Initial)<br>b. Department<br>c. Non-Host Inst. | (4) USAGE FACTOR: Resource Technology a/ | (4) USAGE FACTOR: CPU MIN USED | (4) USAGE FACTOR: Resource Staff Hours | (5) BRTP Fund Allocation |
|---|---|---|---|---|---|---|---|
| BIONET Training Program<br>(FASEB, Rutgers/Waksman, Nature update in Molecular Biology, Biotech '85, Miami Mid-Winter Symposia) . | 9 | 40,42<br>68 | a.<br>Mansfield, Elaine<br>Lawler, MaryJo<br>Azhir, Ari<br>Kedes, Laurence<br>Brutlag, Douglas<br>Smith, Dennis H.<br><br>b.<br>IntelliGenetics | DEC-2060<br>"<br>"<br>"<br>"<br>" | 447<br>68<br>2279<br>71<br>276<br>22 | 151<br>255<br>203<br>40<br>40<br>47 | 8 |
| CUMULATIVE TOTALS: No. subprojects | 1 | | | | 3163 | 736 | $61,348 |

a/ Identify Resource Technologies Used. b/ Give Hours Resource Technologies Used. See Instructions, page ii.

## II.B. Books, Papers and Abstracts

We report the publications by members of the BIONET scientific community on a version of the special form as generated by our local database management system. We report only the category of Collaborative Research and Service. There have been no publications in the Core Research program. The published materials used to support our Training program (the *Introduction to BIONET*, the *BIONET Reference Manual* and the *BIONET Training Manual*) have been described before and thus are not reported separately.

Part II, Section B
INSTITUTION: Intelligenetics

Award Number 1U41RR01685-02
REPORT PERIOD: March 1, 1986 to February 28, 1986

COLLABORATIVE RESEARCH

* Vogt, M., Haggblom, C., Swift, S., Haas, M., Envelope gene and long terminal repeat determine the different biological properties of Rauscher, Friend, and Moloney Mink Cell Focus-inducing viruses, J. Virol. 55, 184-192, 1985

* Upton, C, McFadden, G., DNA Sequence Homology between the Terminal Inverted Repeats of Shope Fibroma Virus and an Endogenous Cellular Plasmid Species, Molecular and Cellular Biology, Jan. 1986

* Andersen, R.D., Birren, B.W., Taplitz, S.J., Herschman, H.R., The Rat Metallothionein-l Structural Gene and Three Pseudogenes, One of Which Contains 5'-Regulatory Sequences, Molecular and Cellular Biology, 1985

* Andersen, R.D., Taplitz, S.J., Briston, G., Herschman, H.R., Rat Metallothionein Multigene Family, in Proceedings of the Second International Meeting on Metallothionein and Other Low Molecular Weight Metal-binding Proteins, Birkhauser Verlag, Boston, Aug. 21, 1985

Katzen, A.L., Kornber, T.B., Bishop, J.M., Isolation of the proto-oncogene c-myb from Drosophila melanogaster, Cell 41: 449-456, June 1985

Simon, M., Drees, B., Kornberg, T., Bishop, J.M., The Nucleotide sequence and tissues specific expression of Drosophila c-src, Cell, Oct 1985

* Lohe, Allan R., Brutlag, Douglas L., Multiplicity of Satellite DNA Sequences in Drosphila Melanogaster, PNAS, 1985

* Calhoun, David, Bishop, David T., Bernstein, Harold S., et. al., Fabry Disease: Isolation of a cDNA clone encoding human alpha-galactosidase A, PNAS, 1985

* Cooke, N.E., David, E.V., Serum Vitamin D Binding Protein is a Third Member of the Albumin and alpha-fetoprotein gene family, J. Clin. Investig, 1985

* Glaichenhaus, N., Leopold, P., Cuzin, F., et. al., Changes in the expression of cellular genes inn cells immortalized or transformed by polyoma virus, Cancer Cells, Cold Springs Harbor Laboratory, 1985

Biggs, J., Searles, L.L., Greenleaf, A.L., Structure of the Eukaryotic Transcription Apparatus: Features of the Gene for the Largest Subunit of Drosophila RNA Polymerase II, Cell 42: 611-621, Sept. 1985

* Hamori, Eugene, Novel DNA Sequence Representations, Nature 314: 585, 1985

* Nakauchi, H., Nolan, G.P., Herzenberg, L.A., et. al., Molecular cloning of Lyt-2, a membrane glycoprotein marking a subset of mouse T lymphocytes: Molecular homology to its human counterpart, Leu-2/T8, and to immunoglobulin variable regions, PNAS USA 82: 5126-5130, 1985

* Hogness, D.S., et. al., Regulation and Products of the Ubx Domain of the Bithorax Comples, Cold Springs Harbor Symposia Vol L, 1985

* Allison, L.A., Moyle, M., Shales, M., Ingels, C.J., Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases, Cell 42: 599-610, 1985

* James, D., Leffak, I.M., Polarity of Replication Through the Avian Alpha-Golbin Locus, Mol. Cell. Biol., 1985

* Singer, P.A., Oshima, R.G., Molecular Cloning and Characterization of the Endo B Cytokertin Expressed in the Preimplantation Mouse Embroyos, J. Biol. Chem., 1986

* Knott, T.J., Rall, S.C., Scotts, J., et.al., Human Apolipoprotein B. Structure of Carboxyl-Terminus

Domains, Sites of Gene Expression, and Chromosomal Location, Science, 1985

* Regier, J.C., Pacholski, P., PNAS USA 82: 6035-6039, 1985

* Robinson, H.L., Miles, B.D., Avian leukosis virus-induced osteopetrosis is associated with the persistent synthesis of viral DNA, Virology 141: 130-143, 1985

* Miles, B.D., Robinson, H.L., High frequency transduction of c-erB in avian leukosis virus-induced erythroblastosis, J. Virology 54: 295-305, 1985

* Shank, P.R., Schatz, L.M., Robinson, H.L., et. al., Sequences in the gag-pol-5' env region of avian leukosis viruses confer the ability to induce osteopetrosis, Virology 145: 94-104, 1985

* Robinson, H.L., Jensen, L., Coffin, J.M., Sequences outside of the LTR determine the lymphomogenic potential of Rous assoicated virus-1, J. Virology 55: 752-759, 1985

* Robinson, H.L., Avian leukosis viruses as vectors for the development of vaccines, Proceedings of 34th Annual National Breeders Roundtable: 64-92, 1985

* Robinson, H.L., Gagnon, G.C., Patterns of proviral insertion and deletion in avian leukosis virus induced lymphhomas, J. Virology, 1985

* Miller, R.H., Robinson, W.S., Common Evolutionary Origin of Hepatitis B Virus and Retroviruses, PNAS, 1985

* Schumacher, M., Camp, S., Taylor, P., et. al., Primary Structure of Torpedo californica Acetylcholinesterase Deduced from dDNA Sequence, Nature, 1985

* Timmerman, K.P., Tu, D., Compete sequences of IS3, N A R 13: 2127-2139, 1985

* Kemper, B., Molecular Biology of Parathyroid Hormone, Critical Reviews in BioChem, 1986

* Maratea, D., Young, K., Young, R., Deletion and Fusion Analysis of the Phi 0X174 E Lysis Gene., Gene, 1985

* Machida, C.A., Bestwich, R.K., Kabat, D., A Weakly Pathogenic Mutant of Rauscher Spleen Focus-Forming Virus Has Lost the Carboxyol-Terminal Membrane Anchor of Its Envelope Glycoprotein, J. Virol. 53: 990-993, 1985

* Machida, C.A., Bestwick, R.K., Boswell, B.A., Kabat, D., Role of a Membrane Glycoprotein in Friend Virus-Induced Erythroleukemia: Studies of Mutant and Revertant Viruses., Virology 144: 158-172, 1985

* Bestwick, R.K., Hankins, W.D., Kabat, D., Roles of Helper and Defective Retroviral Genomes in Murine Erythroleukemia: Studies of Spleen Focus-Forming Virus in the Absence of Helper., J. Virol. 56, 1985

Li, J.-P, Bestwick, R.K., Machida, C.A., Kabat, D., Role of a Membrane Glycoprotein in Friend Viral Erythroleukemia: Nucleotide Sequences of Non-Leukemogenic Mutant and Spontaneous Revertant Viruses., J. Virol. 57, 1986

* Gustafson, T.A., Markham, B.E., Morkin, E. , Analysis of Thyroid Hormone Effects on Myosin Heavy Chain Gene Expression in Cardiac and Soleus Muscles Using a Novel Dot-Blot mRNA Assay., BBRC Vol. 130, No.3 1161:1167, 1985

* Morkin, E., Sheer, D., Gustafson, T.A., et. al., Regulation of Cardiac Myosin Isoenzymes by Thyroid Hormone, UCLA Symposia on Mol., Vol.20, 1985

Black, D.L., Chabot, B., Steitz, J.A., U2 as well as U1 Small Nuclear Ribonucleoproteins are Involved in Pre-Messenger RNA Splicing, Cell 42, 1985

* Smith, D. H., BIONET: National Computer Resource for Molecular Biology, Abstracts, Federation of

COLLABORATIVE RESEARCH

American Socieites for Experimental Biology, Apr 22-25, 1985, Anaheim, CA

* Smith, D.H., BIONET: National Computer Resource for Molecular Biology, Abstracts, International Congress on Computers and Biotechnology, Jan 30-31, 1986, Baltimore, MD

* Smith, D.H., Brutlag, D., Friedland, P., Kedes, L., BIONET: National Computer Resource for Molecular Biology, Nucleic Acids Research, 1986

## II.C. Resource Summary Table

The Resource Summary Table includes the totals from the previous sections under Training and Core Research and Development. The totals for the Collaborative Research and Service categories were arrived at as follows. The Usage Factor, again in cpu minutes, represents the grand total of all use summarized on the previous forms, plus staff use not allocated to the above two categories. The BRTP funds allocated include the remainder of the direct costs estimated for this year (see Section III of the "Application for Continuation Grant"), i.e., those funds not allocated to the above two categories.

The category of Administration/Miscellaneous includes only the Usage Factor of BIONET's share of the cpu time (minutes) for computer facility staff and DEC-2060 system overhead accounts (see Table III-8). No funds for this cpu time or staff time are allocated; such funds are considered part of the support of the user community and are distributed on the basis of cpu time to the categories above, as described previously. The Funds Allocated include only the items of capital equipment purchased during the year.

The category of Down Time includes the sum of scheduled and unscheduled maintenance on the DEC-2060 computer. In the period 12/84 - 11/85, there was a total of 139 hours (8340 cpu minutes) of down time:

- 93 hours (5580 cpu minutes) were scheduled down time, including 69 hours of downtime for the move of the computer over the weekend of August 17-18, 1985; the remainder was scheduled maintenance.
- 46 hours (2760 cpu minutes) of down time were due to unscheduled maintenance.

The down time reported on the Summary Table is BIONET's 50% of the total, or 4,170 cpu minutes. Note that the unscheduled maintenance of 2760 cpu minutes is 0.5% of the total cpu time available for the year. Thus, the DEC-2060 system has been available more than 99% of the time, 24 hours a day, seven days a week. No funds have been allocated to this category.

**PART II, SECTION C RESOURCE SUMMARY TABLE**

AWARD NUMBER 1 U 4 1 R R 0 1 6 8 5 — 0 2

REPORT PERIOD March 1, 1985 to February 28, 198

| RESOURCE COMPONENT | (1) Number Subproject | (2) Number Publications | (3) Number Investigators | (4) USAGE FACTOR Resource a/ Technology | CPU MIN USED | Resource Staff Hrs | (5) BRTP Funds Allocated $ | (6) Resource Fees $ Collected | (7) Other Funds $ |
|---|---|---|---|---|---|---|---|---|---|
| CORE RESEARCH & DEVELOPMENT | 2 | 2 | 3 | DEC 2060 | 627 | 470 | $ 27,543 | - 0 - | - 0 - |
| COLLABORATIVE RESEARCH & SERVICE | 361 | 40 | 1010 | DEC 2060 | 138,095 | 6507 | 540,760 | - 0 - | $97,600 |
| TRAINING | | 0 | 6 | DEC 2060 | 3,163 | 736 | 61,348 | - 0 - | - 0 - |
| ADMINISTRATION/ MISCELLANEOUS | | | | | 17,418 | | 5,000 | | |
| DOWN TIME | | | | | 4,170 | | | | |
| GRAND TOTALS | 363 | 42 | 1029 | | 43,473 | 7713 | $634,551 | - 0 - | $97,600 |



See Instructions, page iv.

# Part III. Narrative Description

## III.A. Summary of Research Progress

This has been a difficult, yet exciting and successful second year for the BIONET Resource.

### Problems and Solutions

The year was made difficult by the unanticipated reduction of this year's grant award at the very end of the first year. Substantial dislocations in our level of staff and planned activities were the result. The Resource has substantial fixed costs in providing computer facilities and telecommunication access to the scientific community. Thus, our final, approved budget had severely reduced funds in categories of personnel, documentation and services, training and other support of Core and Collaborative research activities. In addition, our projections showed that we would exceed our budgeted funds for network communications about eight months into the grant year. These facts led us to take the following steps to preserve the availability of the Resource to the community.

Documentation. We could no longer afford to sell BIONET reference and training manuals to BIONET scientists for the cost of production, because of staff time involved in processing and filling orders. Therefore, we turned the processing of orders for these materials over to IntelliGenetics. We continue to provide, free of charge, the "Introduction to BIONET" for all new PI's.

Computer Networks. We searched for and found an alternative, lower cost vendor for telecommunications network support, and made the transition from Telenet to UNINET in the first part of July, 1985. This switch has allowed us to reduce our costs/access port to the point where, coupled with subscription fees (see next paragraph), we could increase the number of lines available to BIONET scientists outside our immediate area. Further information on this transition and the current network configuration can be found in the letter to BIONET scientists in Appendix I, and in the subsequent Subsection III.A.5, Resource Facilities.

Subscription Fees. Our original proposal had anticipated some level of cost recovery or access charges beginning in the third year of our grant. However, the Year 2 budget cut forced us to impose an access charge, with very short notice for the BIONET community. After negotiations with the NIH Division of Research Grants, we were able to impose a subscription fee, recognizing the monies collected as telecommunications access charges that were not to be regarded as grant-related income (as grant-related income, the monies collected would be used to reduce the total grant award). The letter announcing the NIH view of the access charges is attached to Section III of our "Application for Continuation Grant". A discussion of the fee was sent to all BIONET users (see Appendix I). New BIONET applicants receive a revised version of that text with their application forms. The fee was set at $400 per PI per year, for reasons summarized in Appendix II.

Subsequently, we announced a new category of membership in the BIONET Resource, Class IV, for those who wish to use only the communications aspects of the Resource, electronic mail and electronic bulletin boards. The fee for this method of access is $100 per PI per year. The letter announcing Class IV access, sent to all persons who had not contacted us about payment of subscription fees, is given in Appendix III. A version of this letter was also posted on the BIONET-NEWS bulletin board on BIONET. Information on Class IV access is now part of our standard application package.

## A Successful Year

Despite our fiscal and administrative difficulties, we have been able to accomplish a great deal this year. These accomplishments are described in detail in the subsequent discussion, but here, in brief, are some notable successes:

- We have approved 560 PI's for access to the Resource, representing about 1800 individual scientists.
- Over 50 publications have appeared or are in press, where the facilities of the BIONET Resource played a key role in obtaining results. This is particularly notable in that more than half of the community has had access for only a few months.
- The computer facility is under new management; the facility was moved to our new offices over one weekend, and was available again to the community with less than three days total downtime. Facilities have been improved substantially.
- The imposition of subscription fees has been successful. Not only have they generated enough support to solve our fiscal problems, the impact on the Resource and its community has been relatively minor. With very few exceptions, those choosing not to pay were not significant users of the Resource.
- We have been able to make significant progress in our Core and Collaborative research programs, despite the budget cuts. Two major Core projects have been begun, our BIONET Satellite program, and investigation of hardware solutions to the problem of biological sequence database search, retrieval and analysis.
- Our reapplication forms for another year of access for PI's have led to many words of praise for the Resource and its crucial role in the PIs' research. A few PI's had significant complaints, and we are working with those PI's on their problems.

### III.A.1. Service

The Service component of the BIONET Resource includes primarily Class I investigators who wish to use the BIONET facilities to support their research. It also includes Class III investigators, a classification reserved for persons responsible for Department, School, or Campus-wide computer facilities who wish to provide information about or access to BIONET to the local community they serve. There are currently three investigators who have been granted this status, K. Manly at Roswell Park, P. Vitek at the Imperial Cancer Research Foundation, and J. Claverie at Institut Pasteur.

More recently, we have established a new class of BIONET participants, Class IV. This class is reserved for those investigators who wish to take advantage of the communication components of BIONET, but do not wish to run the analysis programs. This policy should address desires for this level of access expressed at our March, 1985 NAC meeting. There will be a telecommunications-related subscription fee for this level of access as well, and we have set this fee at $100. We have already informed all PI's who chose not to pay the full subscription fee about this policy (see letter in Appendix III). Interestingly, none have chosen to accept this option. We also have sent out a bulletin on the system to inform all current subscribers. We now include this option as part of our standard information and application packet sent out to potential BIONET scientists. Only one new applicant has so far requested this status.

### III.A.1.a. Scientific Consulting - Class I, III and IV Support

An important function of the BIONET consultants is to provide support to BIONET users. This support takes the form of answering questions and making available information about the BIONET Resource. Questions come by phone, by electronic mail, and by terminal links from investigators to staff. A survey of the past year's phone, mail messages, and terminal links from BIONET Resource scientists gives a good overall indication of the usage of the different aspects of the Resource.

A total of 2043 inquiries were tallied from November 1, 1984 to November 13, 1985. The data were divided into five categories by subject. These categories are programs (Core and Contributed Libraries, electronic communication), TOPS20 system, personal computers and personal computer software, telecommunications, and BIONET administration. The distribution among these categories of questions is shown in Table III-1).

**Table III-1:** Summary of Distribution of Questions.

| Category | Number of Inquiries | Percent of Total Inquiries |
|---|---|---|
| Programs | 745 | 36 |
| TOPS20 System | 423 | 21 |
| PC and PC software | 313 | 16 |
| Telecommunications | 287 | 14 |
| BIONET Administration | 275 | 13 |
| TOTALS | 2043 | 100 |

The first category in Table III-1, Programs, has been further subdivided into scientific categories of questions, as summarized in Table III-2). The first column indicates the category. The second column shows the total number of questions received in each category. The third column shows the number of

questions received as a percent of the total questions on programs. The last column shows the number of questions in each category as a percent of the total of 2043 questions tallied above.

**Table III-2:** Breakdown of Program Questions by Function

| Scientific Category | Number of Program Inquiries | Percent of Program Inquiries | Percent of Total Inquiries |
|---|---|---|---|
| Sequence and Gel Data Entry & Manipulation | 104 | 14 | 5 |
| Consensus Sequence Determination | 58 | 8 | 3 |
| Protein and DNA sequence Manipulation | 199 | 27 | 10 |
| Databases, Database Searches and Sequence Alignment | 259 | 35 | 12 |
| Experiment Planning | 19 | 2 | 1 |
| Communication with the BIONET Community | 74 | 10 | 4 |
| Other | 32 | 4 | 1 |
| TOTALS | 745 | 100 | 36 |

As shown in Table III-2, there were 745 requests for information concerning programs offered by the BIONET Resource; these questions comprised 36 percent of the total requests for information. Questions about database access and sequence comparison represented more that 1/3 of the total program inquiries. Questions about biological sequence manipulation, for example, restriction mapping, alignment, translation, codon frequencies, represented 27% of the total. Sequence entry, determination of consensus sequences, and electronic communication make up the bulk of the rest of the questions.

Returning to Table III-1, questions about the TOPS20 system represented 21% of the total. These questions concerned accessing programs, manipulating files and directories, using the text editors, and controlling output on the terminal.

A wide variety of personal computers (PC's) and PC software is in use by BIONET subscribers. The most frequently asked questions concerned IBM PCs and Apples and their related software. The majority of information requested concerned the transfer of files to and from the user's computer and the BIONET computer. There were a total of 313 questions, comprising 16 percent of the total, on this topic.

The telecommunication category includes phone and mail correspondence concerning the quality of communications between remote users and BIONET, and the procedures involved in connecting to the BIONET computer. These questions made up 14 percent of the total with 287 inquiries.

The final category of inquiries concerns BIONET Administration. This category does not include all the correspondence as the majority of questions were directed to the BIONET administrator, and not to the BIONET consultants. There were 275 requests for information. These inquiries covered requests for and complaints about the manuals, account or application status, adding or deleting users, and possibilities for training sessions. This category comprised 13 percent of the total.

Our data indicate that the majority of users have needed assistance on use of the Core Library software to solve specific scientific problems. As the community has matured in its use of the software, the questions have become deeper, indicating a real awareness of the basics of using computers to solve such problems. The large number of questions about the TOPS-20 system, telecommunications, and PC access to the Resource reveal that the community is distributing its computational tasks. PC's are used to collect sequence and other data, and to up load this information to BIONET. Results obtained from analysis on BIONET are then down loaded to the PC's, where they are used to guide new experiments and eventually used in preparation of scientific publications.

### III.A.1.b. Scientific Case Studies Using BIONET

In just over a year of user access, BIONET has grown to service almost 1500 researchers in the field of molecular biology. In one way, BIONET's contribution to research can be inferred by the number of subscribers. However, another way to demonstrate BIONET's current and potential efficacy is through the examination of the work of a few researchers who have submitted research findings for publication and had used BIONET in the process of discovering these findings.

Using information obtained through our reapplication procedures (see below) about BIONET use and resulting publications, we have selected three examples which we present as brief case studies: "Common Evolutionary Origin of Hepatitis B Virus and Retroviruses," PNAS, in press (1986), by R. Miller and W.S. Robinson (Stanford University); "DNA Sequence Homology between the Terminal Inverted Repeats of Shope Fibroma Virus and an Endogenous Cellular Plasmid Species,"Mol. Cell Bio., in press (1986), by C. Upton and G. McFadden (University of Alberta); and "Molecular Cloning and Characterization of the Endo B Cytokeratin Expressed in Preimplantation Mouse Embryos,"J. Biol. Chem., in press (1986), by P. Singer, K. Trevor, and R. Oshima (La Jolla Cancer Research Foundation).

"Common Evolutionary Origin of Hepatitis B Virus and Retroviruses"

Dr. William Robinson, a Professor at Stanford University, was among the first researchers to join

BIONET. Since the acceptance of his application on October 31, 1984, he and Roger Miller, a post doctoral fellow, have used over 64.5 CPU hours in 540 connect hours on BIONET in their research on the molecular structure of Hepatitis B virus.

Initially, Dr. Robinson and Dr. Miller chose to study the secondary structure of the origin of replication of known hepadna viruses. In examining stable palindromes near the origin of replication, Drs. Robinson and Miller discovered by computer manipulation that the regions flanking these palindromes were highly conserved. In an effort to substantiate these findings, they performed several global searches through the Genbank and EMBL databases and found that not only were these regions conserved across hepadna viruses, but they were also present in type C retroviruses. As a result of further analyses in the lab, Dr. Robinson and Dr. Miller gathered additional evidence which led them to suggest that HBV and retroviruses have a common evolutionary origin. The authors have stated that without the use of computer analysis software and access to a complete and up-to-date database, the question of the genetic relatedness of the hepadna virus and retrovirus families would not have been raised.

Although BIONET is not the only source of DNA and protein analysis programs, it was the primary resource for Dr. Robinson and Dr. Miller. The investigators retrieved the hepadna virus sequences from the Genbank database using the QUEST program. They performed the initial DNA homology and palindrome analyses with the SEARCH function of the SEQ program, and employed other SEQ commands to examine base-composition and to translate sequences. Drs. Robinson and Miller discovered that the regions were conserved in 27 viral DNA sequences by searching over the Genbank and EMBL databases using the IFIND program. Additional searching using IFIND and the SEARCH function of PEP demonstrated a high degree of homology between the HBV core protein and the retroviral P30 gag nucleocapsid protein. The investigators also used the PEP program for open reading frame analysis, hydropathicity plots and secondary structure prediction. Dr. Miller attributed their use of BIONET as a resource for their DNA and protein analysis programs to the system's comprehensive suite of software and its inexpensive and easy access.

When asked which components of the resource he thought needed attention, Dr. Miller felt the documentation was a problem for computer naive users. We feel this shortcoming will be addressed with the implementation of more training programs and the addition of an example manual for new users (see Subsection III.A.4). Dr. Miller was, however, particularly positive about the electronic mail and bulletin board facilities on BIONET. As a result of interactions on BIONET, the Robinson lab traded unpublished hepadna virus sequences with several other labs on the system. In addition, Roger Miller actively participated in the community bulletin boards. For the benefit of other BIONET users, Dr. Miller reviewed and documented how to use Michael Zuker's program for predicting RNA secondary structures, BIOFLD. He posted his review on both the BIONET-NEWS and the CONTRIBUTED-

SOFTWARE bulletin boards, and it has proven to be one of the most widely read and used bulletins (see Subsection III.A.2). It is interesting to note that Roger Miller was a novice computer user before joining BIONET. He had previously used a personal computer for only a month.

"Molecular Cloning and Characterization of the Endo B Cytokeratin Expressed in Preimplantation Mouse Embryos"

Dr. Robert Oshima of the La Jolla Cancer Research Foundation in La Jolla California joined BIONET on December 31, 1984. Since that time, he and a Post Doctoral Fellow in his lab, Phillip Singer, have used over 9 CPU hours in 190 connect hours on BIONET in their research of Endo B filament proteins.

Much of the research for this paper had been done prior to the lab's BIONET access, but Dr. Singer used BIONET extensively for the confirmation of their research results. Having already constructed a cDNA clone (Endo B) from parietal endodermal mRNA, Drs. Oshima and Singer characterized their cDNA clone through analyses of amino acid composition of Endo B, open reading frame positioning, and homology comparisons against other known filament proteins and type I keratins.

Dr. Oshima and Dr. Singer used the QUEST and IFIND programs for retrieving known type I keratin sequences and aligning those with the Endo B protein. These comparisons, along with additional searches using the PEP program, confirmed the investigators' findings that the Endo B protein had a 54-68% homology with type I keratins. Dr. Singer also used PEP to further confirm their characterization through hydropathicity and secondary structure analyses. A final search over the Genbank and EMBL databases revealed approximately three to five genes in the mouse genome that showed homology to the Endo B cDNA.

Unlike the Robinson lab, Dr. Oshima had additional sources for computer analysis programs: Roger Staden's nucleic acid analysis programs on the U.C.S.D. computer and the PCS and Schwindigger programs for the IBM personal computer. Because he had prior experience with analysis software, Dr. Singer used the programs in a more sophisticated way, as reflected by his questions to the BIONET Scientific Consultant. He felt that the load on BIONET was unfavorably high during midday hours, but has since resorted to doing his searches in batch to alleviate that problem. However, after doing much of the sequence assembly for sequences reported in the above paper by hand, the Oshima lab now uses BIONET extensively, specifically the GEL program. Because Dr. Singer feels it is useful and convenient to use one set of compatible software when initiating a new research project, he and other members of the lab are using BIONET for the cloning and characterization of an analogous human keratin. They expect this paper to be completed before the end of the year.

"DNA Sequence Homology between the Terminal Inverted Repeats of Shope Fibroma Virus and an Endogenous Cellular Plasmid Species"

Grant McFadden, an Assistant Professor at the University of Alberta, was accepted to BIONET on November 30, 1984. This group, particularly Chris Upton, a post doctoral fellow, has been the heaviest user of BIONET, using the system for over 152 CPU hours in 725 connect hours. Their area of research is the molecular organization of the Shope fibroma gene.

In their paper, Dr. McFadden and Dr. Upton discuss three research findings and suggest how they correlate: the presence of an extrachromosomal autonomous DNA species, its hybridization to Shope fibroma virus, and the exchange of genetic information between host cells and cytoplasmically replicating poxviruses. The investigators used BIONET exclusively for their computer analysis.

The investigators used the GEL program extensively for sequence entry and assembly. Once supplied with the sequence, Dr. Upton used the SEQ program to study the inverted repeat regions of the SFV DNA and to analyze the cytoplasmic DNA molecules through restriction enzyme, base composition, open reading frame, and translation analyses. The investigators' homology comparisons were done using the SEARCH function of the SEQ program. Additional homology searches using IFIND over the Genbank and EMBL databases showed no additional homologous sequences to the inverted repeat region of SFV, but revealed similarity between the extracellular DNA and a family of cellular protease inhibitors.

The McFadden lab is a small lab which, Dr. Upton said, could not afford adequate computer resources without BIONET. His major complaint was the inability to search the entire database, sequences and comments, in one run, a problem that arose from the recent large increases to the databases. Obtaining a solution to this problem is among IntelliGenetics' top priorities. In addition to having access to analytical programs, Dr. Upton is pleased with the opportunity to communicate with other scientists. Like Roger Miller from William Robinson's lab, Dr. Upton had little computer experience prior to BIONET. He has also become very active in the bulletin board communities. Dr. Upton has traded codon usage tables with several other BIONET scientists, and has become one of the community's MacIntosh authorities. Dr. Upton is currently working with an investigator in New York, whom he met through interactions on BIONET, and they are setting up what they call a "personal network" for their collective analysis needs. He foresees using BIONET even more extensively than in the past, especially because the McFadden lab has sequenced 15 to 20KB since he began work in the group.

Our primary service goal at BIONET is to provide a full range of computational tools and a high level of user support. BIONET was also designed to enhance communication among the research community. The results of the above investigators' use of BIONET demonstrates that all Resource components are important in effective use of BIONET for scientific research. All three lab groups stated that they plan to use BIONET even more actively than they had in the past, which was also one of the most frequent comments from the BIONET reapplication forms. In our opinion, we are meeting our service goals and BIONET is becoming established as a useful and effective molecular biology computing resource.

### III.A.1.c. KERMIT Lending Library

The BIONET Resource has been providing access to Kermit software and documentation through a lending library. We have provided this service because of the tremendous number of BIONET scientists with personal computers who wish to access the Resource and require terminal emulation and file transfer. KERMIT provides both.

Kermit is a protocol for transferring sequential files between computers of all sizes over ordinary asynchronous telecommunication lines. Developed at Columbia University, Kermit is non-proprietary, documented, and in wide use. It is presently available for more than 100 different machines and operating systems and additional versions are under development. Because of its wide acceptance, ease of use and minimal cost, we promote the use of Kermit by making it readily available to our users for copying.

We have responded to approximately 70 requests for Kermit software since February, 1985. The lending library provides diskettes and documentation for duplication for the Apple II, IBM-PC and its lookalikes, Macintosh, and the TRS-80. The user can also down load the Kermit files directly from the DEC 2060 for other systems. As the popularity of Kermit increases in the BIONET community, we expect a heavier demand for this software. Some users may prefer to purchase Kermit diskettes and documentation (at a minimal cost) from outside vendors rather than duplicate and return our copies. Columbia University is now selling diskettes for some systems at $10.00 a copy and there are other companies selling diskettes for other systems. As more information becomes available, we will pass it along to our users.

### III.A.1.d. Subscription Fee and Its Effects

The rationale and policies for a subscription fee for telecommunication access to BIONET were introduced at the beginning of Section III.A. Class II and III users are exempt, as are foreign users who pay their own telecommunication costs.

There was no way to determine in advance the effects of this fee on the community. Although the community was polled by electronic mail, few replied and only one or two had serious concerns. From the beginning, we established the policy that no one would be prevented from accessing the Resource because of inability to pay. However, we have also requested that those who cannot pay give us a target date (e.g., for pending grant renewals) when payment might be possible.

We have looked at statistics on the community and its use of BIONET to determine the effects of the fee. The pattern of use of the system as measured by numbers of cpu minutes and connect hours consumed is discussed in detail in Paragraph III.A.5.b. Briefly, we saw no significant decrease in use of the system during the initial period when the fee was imposed that we could ascribe to the fee itself. Summer vacations are a better explanation for a slight decrease in use, because the use has again begun to climb dramatically.

More revealing are the effects on the numbers of PI's accessing the Resource. As summarized in Table II-1, of the 537 Class I PI's who have been accepted, only 27 refused, in writing, to pay the fee. However, nearly 200 PI's have never contacted us at all, so their accounts have been frozen awaiting further word. This was actually not surprising, because only about 65% of the PI's accepted ever used the system even before the fee was imposed. Those who declined and those who have not contacted us have been, almost without exception, very minor users of the system if they logged in at all. Almost 300 PI's have paid the fee or have made arrangements with us that the fee will be paid in the near future. 28 are exempt (Class II, foreign, NAC), giving us a total of 322 active PI groups as of December 1, 1985. Thus, the active community represents about 60% of the total users accepted over the last 15 months. Some percentage of the 40% who are not active users were probably deterred by the fee, but the precise number is difficult to determine. Those who remain find the system useful for their research, and many have told us that the fee is minimal compared to the services available on BIONET.

### III.A.1.e. Reapplication Procedure

One of the original policies of BIONET was that each PI should be required to reapply each year. We have devised a simple procedure that collects information from each PI which we are required to report as part of our Annual Report to the NIH, especially a publication list and rationale for use of BIONET. At the same time, we ask for updates on information supplied previously to determine eligibility for access. We also ask for changes in status of funding, changes in the PI's user group, and so forth. At the request of Dr. Lederberg of our National Advisory Committee, Each PI is again required to supply his or her signature and that of a responsible administrative official,

We have followed this procedure for the preparation of this Report. The reapplication form is shown in Appendix IV. The publication list in Section II.B was compiled from the responses. Along with collection of factual information, we also asked for information on how BIONET was used and for comments on the Resource.

**Use of the Resource, by Scientific Category.** The data collected from the reapplication forms on how BIONET was used is shown in Table III-3. Although not all scientists indicated their patterns of use, and some had not had access long enough for them to comment, the patterns of reported use are almost exactly the same as the pattern of queries to the Scientific Consultants (compare to Table III-2).

**Comments on the BIONET Resource.** 34% of the reapplication forms had some type of comment about the Resource. The majority of comments were favorable; many users felt that the programs and support were excellent. The most frequently cited complaints were about the program documentation and the telecommunication networks. The complaints about the telecommunications centered around slow response time, unexpected disconnections, and difficulty in accessing the BIONET computer. The

**Table III-3:** Breakdown of Program Use, by Category

| Scientific Category | Program Usage | Percent Of Total |
|---|---|---|
| Sequence and Gel Data Entry and Manipulation | 35 | 10 |
| Consensus Sequence Determination | 26 | 7 |
| Protein and DNA sequence Manipulation | 121 | 33 |
| Database Searches/Sequence Alignment | 123 | 34 |
| Experiment Planning | 9 | 2 |
| Communication with the BIONET community | 32 | 9 |
| Other | 17 | 5 |
| TOTALS | 363 | 100 |

reapplicants also included many useful suggestions on programs or features that they would like to see included in the BIONET Resource. The comments from the reapplicants indicate that the users are actively participating in BIONET and are willing to help us better serve the BIONET community.

The specific complaints and suggestions are already being incorporated into our plans for next year. For example, complaints about the manuals (primarily that they are not designed for the naive user), have already prompted new directions in our Training program (see Subsection III.A.4). We have solved most of the network problems, but this is a continuing challenge. Limitations to access have been alleviated somewhat, as we have gone from eight UNINET network ports to twelve.

The suggestions about new programs give us some specific guidance on areas to which we are directing our Core and Collaborative Research efforts. Our Core project on hardware serial search machines is prompted by complaints about the cpu time required to search the growing databases (see Core Research, Subsection III.A.3). Requests have been made for programs for multiple sequence alignment and methods for investigating evolutionary relationships among biological sequences. The former is already being addressed (see Collaborative Research, Subsection III.A.2). The latter has resulted in us approaching scientists with programs for this application as potential collaborators on BIONET.